Genomics 101 (2013) 221–228

Contents lists available at SciVerse ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Bioinformatic survey for new physiological substrates of Cyclin-dependent kinase 5

Daniel A. Bórquez [a], Cristina Olmos [a], Sebastián Álvarez [a], Alex Di Genova [b], Alejandro Maass [b], Christian González-Billault [a,*]

[a] Laboratory of Cell and Neuronal Dynamics (CeNeDyn), Department of Biology and Institute for Cell Dynamics and Biotechnology (ICDB), Universidad de Chile, Chile
[b] Center for Mathematical Modelling (CMM) and Center for Genome Regulation (CGR-Fondap 15090007), Universidad de Chile, Chile

ARTICLE INFO

Article history:
Received 9 July 2012
Accepted 26 January 2013
Available online 4 February 2013

Keywords:
Cyclin-dependent kinase 5
Position specific scoring matrix
Protein phosphorylation
Neurotransmission

ABSTRACT

Cyclin-dependent kinase 5 (Cdk5) is a proline-directed serine/threonine kinase predominantly active in the nervous system where it regulates several processes such as neuronal migration, cytoskeletal dynamics, axonal guidance, and neurotransmission.

We constructed a position specific scoring matrix (PSSM) based on a dataset of sites shown to be phosphorylated both *in vivo* and *in vitro* by Cdk5. This dataset was curated manually through an exhaustive search of published experimental data. We then used this PSSM to perform a search in the mouse proteome through Scansite, a web-based tool for matching sequence patterns in large databases. Considering a stringent cut-off score of 0.5, we identified 354 new putative sites present in 291 proteins. In order to assess the robustness of our results, ten random subsets (of 80 sites each) of the original dataset were used to construct new PSSMs, which were then used as input for a new Scansite search, leading to the recovery of 81% of the 354 sites by at least 5 PSSMs.

In order to reduce the number of false positives in our sequence-based approach, we evaluated which of these predicted sites were phosphorylated *in vivo* as determined by multiple phosphoproteomics studies carried out through mass spectrometry and available in the PhosphoSitePlus database. This step resulted in a very promising list of 132 putative phosphorylation sites for Cdk5, of which, 51 are specifically phosphorylated in brain tissue, and some are involved in functions regulated by Cdk5 such as axonal growth, synaptic plasticity and neurotransmission. Other phosphorylation sites in our list suggest that Cdk5 might regulate processes through mechanisms not previously recognized such as the control of mRNA splicing.

1. Introduction

Cdk5 is an atypical member of the Cdk family, as its activity is regulated through the binding of two non-cyclin activators which are present in postmitotic neurons, namely p35 (Cdk5Rr1) and p39 (Cdk5r2). Additionally, Cdk5 is not directly involved in cell cycle regulation, but controls multiple aspects of neuronal function such as axonal elongation, neuronal migration and synaptic transmission [1]. Recently it has been shown that Cdk5 participates in other physiological functions that occur outside of the nervous system (reviewed in [2,3]).

Cdks are among the most promiscuous kinases, being able to phosphorylate hundreds of substrate proteins [4]. In fact, a global analysis of yeast has found that Cdk1 can phosphorylate between 181 and 308 target proteins [5,6]. Kinase diversity and pathway overlap precludes the conduction of global studies in the higher eukarya proteomes, where the search for new kinase-specific substrates has been mainly performed through computational methods. Such strategies present many disadvantages, mainly the large proportion of false positives in the results obtained. Until now, only one bioinformatic study has been reported to predict phosphorylation sites for a Cdk in the yeast proteome [7]. Moreover and to our knowledge few studies have succeeded in experimentally verifying computationally predicted sites [8–10].

In this work, we performed the first systematic computational search for putative phosphorylation sites for a Cdk family member (Cdk5) in a mammalian proteome (*Mus musculus*). One of the main pitfalls for the bioinformatic determination of novel kinase-specific phosphorylation sites at a proteome-wide scale is the selection of an appropriate classification method to distinguish between true and false substrates. Currently the most widely accepted methods to achieve this task are position-specific scoring matrices (PSSMs) [11], hidden markov models (HMMs) [12], artificial neural networks (ANNs) [13] and support vector machines (SVMs) [14].

PSSMs are a relatively simple method, which have been successfully used for the prediction of novel phosphorylation sites in proteome-wide studies [11,15]. PSSMs adequately consider the modular nature of the kinase-substrate interactions, where the catalytic domain of the kinase establishes direct contact with a linear motif of the substrate. Such a modular characteristic explains the tendency of kinases to phosphorylate a particular sequence motif, which in the case of Cdk5

* Corresponding author. Fax: +56 2 22712983.
E-mail address: chrgonza@uchile.cl (C. González-Billault).

0888-7543/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2013.01.003

is a serine (Ser) or threonine (Thr) followed by a proline (Pro) residue. As *bona fide* phosphorylation sites in a proteome are estimated to be less than 10% of all potentially phosphorylated residues, even at a low false positive rate, the expected number of false positive candidates precludes experimental validation [16].

Considering these limitations, we developed a strategy that considers a protein as a good candidate to be phosphorylated by Cdk5 if the predicted site derived from our bioinformatic analyses has been actually identified as phosphorylated *in vivo* in studies using mass spectrometry. Phosphoproteomic studies describe tens of thousands of *in vivo* phosphorylated sites; however, most of those sites are still orphan for the kinase responsible for this modification. As phosphoproteomics studies increase in number, size and depth, we approach a point at which all the phosphorylated sites in a proteome will be known [17], allowing a significant increase in the reliability of database-wide studies. In this paper we used the sequences known to be phosphorylated by Cdk5 to predict new sites modified by this kinase. These putative sites were then evaluated for their *in vivo* phosphorylation status, in order to improve the reliability of the prediction of new substrates for Cdk5 phosphorylation in the mouse proteome.

2. Results and discussion

2.1. Characterization of known substrates of Cdk5

Table S1 lists the known substrates for Cdk5 and their respective phosphorylation sites as derived from published records (see Methods). Sequence analysis reveals that only 49% of the substrates contain the canonical consensus motif, (Ser/Thr)-Pro-X-(Arg/Lys/His). There is high prevalence of suboptimal motifs containing only the minimal (Ser/Thr)-Pro dipeptide motif which is not exclusive for Cdk5 substrates since previous studies have shown a similar behavior for Cdk1 and Cdk2 substrates, where 39–67% of the phosphorylation sites match only to the minimal proline-directed consensus motif [18,19]. Therefore, large representation of non-consensus motifs makes it especially difficult to predict phosphorylation sites for these kinases. To overcome this, we focused our effort on developing a PSSM (Table 1), which provides a better description of the phosphorylation sites, including sequences distant from the consensus motif. PSSMs have been previously implemented for the search of novel substrates for Cdc28 (Cdk1) in yeast [7], where the construction of the PSSM was based on kinetic data from *in vitro* phosphorylation of a peptide library. Unlike this reported approach, our PSSM was based only on sites shown to be directly and physiologically phosphorylated by Cdk5, which may increase the accuracy of prediction, since the selectivity of protein kinases *in vivo* is higher than *in vitro* [20].

In addition to the sequence motifs found in substrates, other features have been implemented to improve the reliability of the phosphorylation site descriptor. Most commonly used are the structural disorder degree [21] and the clustering of phosphorylation sites in protein sequences [22]. We evaluated the collected dataset of known Cdk5 phosphorylation sites with a disorder prediction tool, showing that the majority of the sites (76%) were located in areas of structural disorder (PrDOS score > 0.5, Fig. 1). However, the analysis of the unphosphorylated (Ser/Thr)-Pro motifs showed the same distribution, indicating a general preference for enrichment of the (Ser/Thr)-Pro dipeptide in areas of intrinsic structural disorder (Fig. 1, red curve). Therefore, in the analyses of Cdk5 phosphorylated sites, structural disorder lacks a predictive value.

2.2. PSSM performance

In order to test the Cdk5 phosphorylation site prediction capacity of our PSSM, we used Scansite to evaluate all Ser and Thr residues in the proteins derived from our bibliographic search (Table S1). Based on our PSSM, Scansite calculated a single numerical value for each 15-residue long sequence centered on a proline preceded by a Ser or Thr. Lower values reflect a better match to the Cdk5 PSSM. Within the Cdk5-phosphorylated proteins, we considered as false positives, all those predicted sites which had not been experimentally shown to be phosphorylated by this kinase, with the exception of those located in transmembrane or extracellular segments of membrane proteins (according to the Swiss-Prot topological annotations as “Extracellular” or “Transmembrane”), which were discarded. This is exemplified by the data collected for the amyloid precursor protein (APP), where Cdk5-dependent phosphorylation has been shown to occur at Thr668 [23]. In addition to the correct identification of this site, our PSSM-mediated analysis predicted one additional site (Thr306) positioned at the extracellular domain, which therefore was not considered as a false positive. Fig. 2A shows the distribution of true positives, true negatives, false positives and false negatives as a function of the Scansite score. The PSSM identified 92.7% of the sites as true positives, but was unable to discriminate properly between these sites and the false positives, especially at higher score values.

Table 1
Amino acid preferences of Cdk5 around the phosphorylation site. In green are highlighted scores that show significant enrichment (log (score) ≥2) and in red those exhibiting depletion (log (score) ≥−2).

		Amino acids																			
		Ala	Cys	Asp	Glu	Phe	Gly	His	Iso	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
Positions	-6	1.06	0.47	0.44	0.90	0.00	1.77	0.41	0.97	1.30	0.42	0.97	0.88	1.34	1.10	1.11	1.50	0.97	1.91	0.00	0.00
	-5	1.20	0.92	1.09	0.89	0.29	0.80	0.80	0.48	1.47	0.63	0.96	1.17	1.49	1.31	1.29	1.36	0.58	0.69	0.00	1.96
	-4	0.60	0.46	1.74	1.19	0.57	0.64	0.00	0.24	0.37	1.05	0.00	2.04	1.32	0.87	0.55	2.85	1.34	0.34	0.00	0.78
	-3	1.35	0.92	1.09	0.89	1.14	0.64	1.21	0.00	0.37	0.95	0.00	1.17	1.49	0.66	2.02	1.11	1.34	0.86	1.68	0.78
	-2	0.90	0.46	0.87	0.59	0.86	0.16	0.00	0.48	1.29	0.53	2.39	0.88	2.65	1.09	1.10	1.61	1.15	1.37	0.00	0.39
	-1	0.75	0.46	0.44	0.44	0.57	0.96	0.80	0.72	1.29	1.47	0.96	0.58	1.16	1.31	1.10	1.86	1.34	0.69	0.00	0.78
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.78	4.80	0.00	0.00	0.00
	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	15.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	1.05	0.46	0.22	0.74	1.14	1.75	0.40	0.72	1.47	0.74	0.00	0.88	1.16	1.75	2.02	0.99	1.15	0.51	0.84	0.39
	3	0.90	0.46	0.22	0.30	0.29	0.48	2.01	0.72	3.86	0.42	0.00	0.29	1.16	0.44	4.59	0.62	0.77	0.51	0.00	0.78
	4	1.05	0.46	1.09	0.89	0.29	1.11	2.41	1.43	0.55	0.53	0.96	1.46	1.16	1.09	1.84	1.24	1.34	0.86	0.84	0.39
	5	1.51	0.46	1.09	0.30	0.57	1.75	1.21	1.48	1.47	0.63	0.48	0.29	0.50	1.53	2.20	1.36	1.34	0.34	0.00	0.78
	6	2.41	0.00	1.31	0.74	0.29	0.64	0.00	0.24	0.73	0.63	1.91	1.17	1.82	1.31	0.92	1.36	0.58	1.20	0.00	0.78
	7	1.20	0.46	1.09	1.19	0.57	0.32	0.40	0.24	0.73	1.16	1.43	0.29	1.99	1.09	1.65	0.87	1.73	0.86	1.68	0.00
	8	2.11	0.46	0.22	0.44	1.14	0.96	0.40	0.48	0.92	0.95	0.96	1.17	1.66	1.97	0.55	1.61	0.77	0.51	0.00	0.78

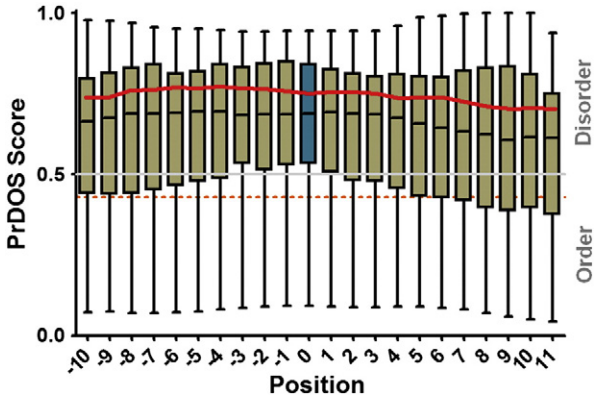

Fig. 1. Structural disorder around Cdk5 phosphorylation sites. Box plots of the PrDOS score around the phosphorylation site (in red). The bottom and the top of the box are the 25th and 75th percentiles respectively; the central band is the median; the *whiskers* extend to the most extreme data points. The red line represents the median of the S/T[P] sites not phosphorylated by Cdk5. The gray horizontal line represents the limit score between order/disorder (PrDOS score = 0.5) and the dashed orange line represents the average score of all amino acids in the Cdk5 substrates.

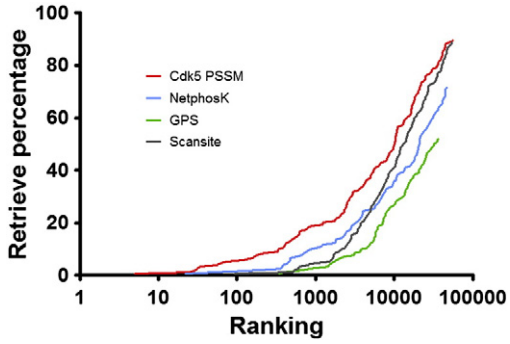

Fig. 3. Comparative performance of different methods for finding Cdk5-phosphorylated sites in the mouse proteome. The performance of a search in the mouse proteome using our custom Cdk5 PSSM in retrieving known Cdk5-dependent phosphorylation sites (Table S1) was compared with the performance of three previously implemented methods for searching kinase-specific phosphorylation sites: Scansite (with the default Cdk5 PSSM), NetPhosK (with a cut-off score of 0.5) and GPS. Our strategy identified a greater percentage of known phosphorylation sites (90%) and assigned to them a comparatively better ranking in the overall results than previously implemented methods.

However, choosing a lower threshold score decreased the number of false positives. Therefore we set up 0.5 as the threshold value, with a false positive percentage of 1%, which is the point where the number of false positives starts to increase exponentially. The compromise between sensitivity and specificity of a classifier is commonly represented by a receiver operating characteristic (ROC) curve (Fig. 2B). The area under our ROC curve (AUC) is a measure of the probability of correctly classifying a Cdk5-phosphorylated residue. In our case, the estimated AUC is 0.7545, indicating a good performance of our predictions.

2.3. Database search

We then used our Cdk5 PSSM to perform a search of the mouse proteome at the Scansite website, producing 400 sites with a lower score than 0.5 (Table S2). These 400 sites were present in 332 proteins, representing 2% of the total proteins in the database. Twenty of these sites had been previously described as Cdk5 substrates. Although this represents a low retrieving rate, our strategy clearly outperforms several previously published methods including NetPhosK [24], GPS [25] and Scansite in a search for Cdk5-dependent phosphorylation sites in the whole mouse proteome (Fig. 3). This is surprising, because NetPhosK and GPS are significantly more sophisticated methods for kinase-specific phosphorylated site prediction. The main pitfall of PSSMs is that they are unable to detect patterns in which combinations of amino acids are relevant. In this line, it was recently found that interpositional dependence on substrate sequence was minimal for various kinases, including one member of the Cdk family (Cdk1) [26], indicating that PSSMs can provide good representations of the phosphorylation sites. Four of the predicted sites are present in Cdk5 substrates but have yet to be reported as phosphorylated by this kinase; therefore, we considered them as false positives. Additionally, 22 of the predicted sites were topologically inaccessible to Cdk5, because they are located in transmembrane or extracellular segments of membrane proteins. Therefore, our search identified 354 sites present in 291 proteins, as new putative phosphorylation sites for Cdk5. To examine the robustness of the new predicted sites, ten different PSSMs were created out of 80 random sites from the *in vivo* and *in vitro* positive dataset to scrutinize the same mouse database using Scansite. This search recovered 286 of the original 354 new putative sites (81%), with each predicted site being present in at least 5 of these shuffled PSSMs (Table S4). Additionally, 60 of these sites were found between the 400 best results for Cdk5 in a search of the mouse proteome by NetPhosK (Table S4), a kinase-specific predictor based on ANNs, which is trained with a different dataset and uses a nine-residue window around the phosphorylation site (unlike the 15 employed in the construction of our PSSM).

Using the PhosphoSitePlus database, we checked whether the 354 new putative sites were actually phosphorylated *in vivo*. Of these, only 130 sites (in 108 proteins) have been reported to be phosphorylated *in vivo* in high-throughput studies (Table S3). These sites represent the most likely candidates to be phosphorylated by Cdk5 in the mouse proteome. Considering that brain tissue displays high Cdk5 activity, we summarized our results on the putative novel Cdk5-

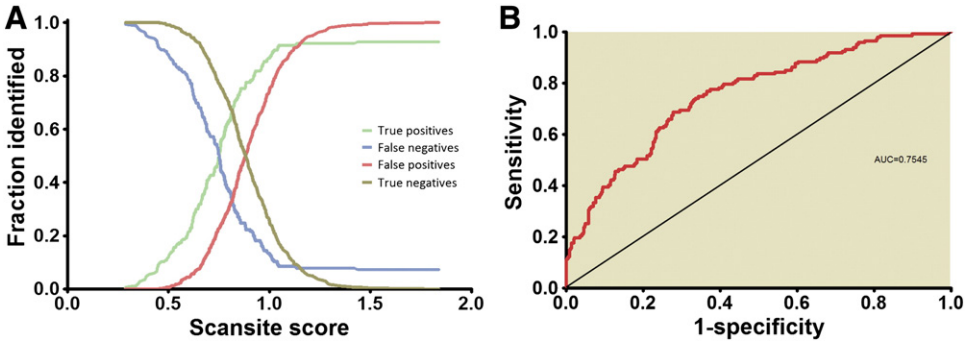

Fig. 2. The Cdk5 PSSM identifies *bona fide* phosphorylation sites in known Cdk5 substrates. A. Plot of the calculated distribution of true positives (green line), false positives (red line), true negatives (brown line) and false negatives (blue line) for a given PSSM score threshold. Note that the false positives were reduced to less than 1% when the cut-off value was set to 0.5. B. ROC curve plotting including AUC value. The diagonal (AUC = 0.5) represents the performance of a random classifier.

Table 2
Putative sites phosphorylated by CDK-5 detected *in vivo* in brain samples.

Protein	Gene name	Site(s)	Commentaries
mRNA processing and splicing			
Protein SON	Son	S1937(a,b)	Acts as a mRNA splicing co-factor. Essential for nuclear speckle organization.
RNA-binding protein with serine-rich domain 1	Rnps1	S27(d)	Component of the active spliceosome. Associated with CDK-11. Phosphorylation by CK-2 in the N-terminal represses its activity.
Serine/arginine repetitive matrix protein 1	Srrm1	S308(b)	Component of the active spliceosome. Involved in numerous pre-mRNA processing events.
Serine/arginine repetitive matrix protein 2	Srrm2	S882(a,b), S1998(a,b,c)	Component of the active spliceosome. *Knockdown* causes defects in cell migration.
Serine/arginine-rich splicing factor 1	Srsf1	S227(b), S238(b)	Phosphorylation in S227 and S238 by Dyrk1A represses the splicing of tau exon 10.
Thyroid hormone receptor-associated protein 3	Thrap3	S55(b,d)	Component of the spliceosome. Mediates mRNA splicing and degradation.
Zinc phosphodiesterase ELAC protein 2	Elac2	S195(a)	Uncharacterized protein. Associated with γ-tubulin complex.
Synaptic transmission			
Activated CDC42 kinase 1	Tnk2	S622(b)	Non-receptor tyrosine and serine/threonine kinase activated downstream of Cdc42. Localizated in presynaptic axons terminals and dendritic spines.
Caskin-1	Caskin1	S1304(f,g)	Scaffold protein that interacts with CASK and cytoplasmic tail of neurexin-1.
Dapper homolog 3	Dact3	S374(a)	Scaffold protein that interacts with Wnt/β-catenin signaling components. Highly expressed in adult cerebral cortex.
Metabotropic glutamate receptor 1	Grm1	S1154(a,b)	G-protein coupled receptor for glutamate.
Microtubule-associated serine/threonine-protein kinase 1	Mast1	S163(a,b)	Postsynaptic protein that interacts with β2-syntrophin.
Nitric oxide synthase, brain	Nos1	S287(a)	Produces nitric oxide dependent on calcium/calmodulin activation.
Paralemmin-1	Palm	S116(a,b)	Essential for dendritic spine maturation.
Phosphatidylinositol 4-kinase type 2-alpha	Pi4k2a	S51(a,b)	Concentrated in Golgi complex and synapses, accounts for most of the phosphotidylinositol-4 kinase activity in the brain.
Protein bassoon	Bsn	S1497(a,c)	Involved in the organization of axonal active zone and neurotransmitter release.
Protein piccolo	Pclo	S4562(a,b,f)	Scaffold protein involved in organization of axonal active zone and synaptic vesicle trafficking.
Protein TANC1	Tanc1	S445(a,b,e)	Postsynaptic scaffold protein required for spine maturation and spatial memory.
Protein TANC2	Tanc2	S1480 (f)	Postsynaptic scaffold protein required for spine maturation. KO is embryonic lethal.
Transcriptional regulation			
AF4/FMR2 family member 4	Aff4	S175(d)	Associated with Cyclin T1/CDK-9. May be have a role in transcriptional regulation and RNA splicing.
B-cell CLL/lymphoma 9-like protein	Bcl9l	S988(a)	Enhanced β-catenin/TCF-mediated transcription.
B-cell lymphoma/leukemia 11A	Bcl11a	S332(a)	Zinc finger transcription factor that regulates *DCC* and *map1b* gene expression.
Cyclin-K	Ccnk	S329(a,b,d)	Regulatory subunit of CDK9, that phosphorylates carboxy-terminal domain of RNA polymerase II to promotes transcriptional elongation.
E1A-binding protein p400	Ep400	S321(a)	Helicase component of Tip60 acetyltransferase multiprotein complex.
Forkhead box protein K1	Foxk1	S406(a,b)	Transcription factor that acts as co-repressor with Sin3.
PC4 and SFRS1-interacting protein	Psip1	S176(a,d,e,f)	A transcriptional activator that binds to promoter elements of heat shock and stress-related genes.
Transcription initiation factor TFIID subunit 9	Taf9	T159(a)	TATA-binding protein (TBP)-associated factor essential for the regulation of RNA polimerase II-mediated transcription.
Zinc finger protein AEBP2	Aebp2	S21(a)	Gli-type zinc finger transcription factor mainly expressed in cells of neural crest origin.
Cytoskeleton dynamics, cell adhesion and vesicular traffic			
Cingulin-like protein 1	Cgnl1	S252(b)	Regulates RhoA and Rac1 activity in tight junctions through interaction with Tiam1 and GEF-H1.
Hepatocyte cell adhesion molecule	Hepacam	S386(f)	Immunoglobulin-like cell adhesion molecule.
Plakophilin-4	Pkp4	S336(b,d,f,g)	Member of a family of armadillo proteins, localizated in desmosomes and adherens junctions. Regulate RhoA activity during cytokinesis.
Pleckstrin homology domain-containing family A member 7	Plekha7	S904(a,b)	Associate microtubules with E-cadherin. Recruiting cingulin-like protein 1 to adherens and tight junctions.
Protein transport protein Sec24A	Sec24a	T227(a)	Regulates ER export by binding to LL signals.
Rootletin	Crocc	S1483(a)	Major component of ciliary rootlet. Contributes to centrosome cohesion.
Scavenger receptor class F member 2	Scarf2	S640(b)	Probable adhesion protein.
VPS10 domain-containing receptor SorCS1	Sorcs1	S1155(b)	Type I transmembrane receptor that mediates intracellular trafficking. Associates with the amyloid precursor protein and regulates Aβ secretion.
Others functions			
Choline-phosphate cytidylyltransferase A	Pcyt1a	S323(a)	Catalyzes the rate-limiting step of phosphatidylcholine synthesis. Ubiquitously expressed.
Choline-phosphate cytidylyltransferase B	Pcyt1b	S323(a)	Catalyzes the rate-limiting step of phosphatidylcholine synthesis. Mainly expressed in the brain. Phosphorylated *in vivo* and *in vitro* by Cdk5.
DNA replication licensing factor MCM2	Mcm2	S27(b,h)	Component of the putative DNA helicase essential for "once for cell cycle" DNA replication initiation.
F-box/WD repeat-containing protein 10	Fbxw10	S942(d)	Substrate-specific adapter for RING uniquitin ligases. Promotes heterochromatin protein (HP)-1α y-1β degradation.
Insulin receptor substrate 2	Irs2	S388(a,b)	Intracellular adapter of insulin and insulin-like growth factor I receptors.
Kelch-like protein 9	Klhl9	S603(b)	Substrate-specific adapter for Cul3 ubiquitin ligase. Promotes Aurora B remotion from chromosomes by ubiquitination and degradation.

Table 2 (*continued*)

Protein	Gene name	Site(s)	Commentaries
Others functions			
Nuclear-interacting partner of ALK	Zc3hc1	S394(d)	Substrate-specific adapter for SCF ubiquitin ligase. Controllers' mitotic entry through Cyclin B1 degradation.
Protein FAM63A	Fam63a	S440(b)	Unknown function.
Protein shisa-6 homolog	Shisa6	S398(a,f)	Unknown function.
Serine/threonine-protein kinase 32C	Stk32c	S18(b,d)	Unknown function.
Sister chromatid cohesion protein PDS5 homolog B	Pds5b	S1166(a,b,d), S1368(a)	Regulatory factor of cohesion complex that maintains sister chromatid linkage until anaphase. Highly expressing in postmitotic neurons.
TBC1 domain family member 13	Tbc1d13	S184(a)	May act as a GAP for Rab GTPases.

(a) Huttlin et al., 2010; (b) Wisniewski et al., 2010; (c) Collins et al., 2005; (d) Tweedie-Cullen et al., 2009; (e) Trinidad et al. 2008; (f) Trinidad et al., 2006; (g) Munton et al., 2007; (h) Ballif et al., 2004.

dependent phosphorylated sites detected by published mass spectrometry studies carried out using brain tissue (Table 2).

2.4. Proteins involved in mRNA processing and splicing

Amongst the putative substrates of Cdk5 identified here, we found proteins that participate in the processing and splicing of mRNA. Although there is no substantial evidence for a role of Cdk5 in these processes, an inhibitory role for Cdk5 in the incorporation of exon 10 in the microtubule associated protein tau has been described [27]. Tau exon 10 splicing is controlled by several proteins, including the serine/arginine-rich splicing factor 1 (Srsf1) (also called ASF/SF2) whose activity is repressed by a Dyrk1a-mediated phosphorylation of three residues (Ser227, Ser234 and Ser238) [28]. Interestingly, two of these sites in ASF/SF2 are identified by our PSSM as candidates for Cdk5 phosphorylation (see Table 2), thus presenting a likely explanation of the role of Cdk5 in the splicing of tau exon 10.

2.5. Proteins involved in neurotransmission

The presynaptic nerve terminal, the main site for neurotransmitter release, is a complex cytomatrix where synaptic vesicle traffic occurs in a continuous cycle of exocytosis, endocytosis and recycling, along with continuous mobilization between various internal vesicles pools [29]. Cdk5 is known to be an active player at presynaptic terminals, as reflected by the multiple targets that localize to this compartment (Fig. 4, red stars). We predict that two of the major scaffold proteins present at the presynaptic cytomatrix, Piccolo (Pclo) and Bassoon (Bsn) could be novel targets for Cdk5 (Fig. 4, orange stars). Additionally, our data suggests that phosphatidylinositol (PI)-4 kinase type IIα (PI4KIIα) (Pi4k2a) could also be a likely substrate. Interestingly, PI4KIIα in conjunction with other known Cdk5 targets such as synaptojanin 1 (Synj1) (Fig. 4, red stars) and PI(4) phosphate 5 kinase type I (PIPKI)-γ (Pip5k1c) (Fig. 4, red stars), regulates PI metabolism at presynaptic terminals (Fig. 4).

Cdk5 acts as a potent suppressor of neurotransmitter release, as its inhibition unlocks the reserve synaptic vesicle pool and accelerates exocytosis kinetics, although the mechanisms involved in this phenomenon remain unknown [30]. Interestingly Piccolo, like Cdk5, is a negative regulator of synaptic vesicle exocytosis, and acts by reducing vesicle translocation from the resting pool to the active zone [31]. Our predicted phosphorylation site on Piccolo (Ser4562) is located between the PDZ domain (residues 4391–4506) and an atypical Ca^{2+}-binding domain (C_2A, residues 4588–4730). The PDZ domain acts as a docking site for Epac2 (Rapgef4), which mediates cAMP-induced exocytosis [32], whereas the low calcium-affinity C_2A domain could act as a presynaptic calcium sensor that mediates the response to repetitive stimulation [33]. Potentially, Cdk5-mediated phosphorylation of Piccolo could regulate the exocytosis events controlled by these two second messengers.

On the other hand, the putative phosphorylation site identified in Bassoon (Ser1497) is located between three adjacent binding sites for the dynein light chains 1 and 2 (DLC1 and DLC2, encoded by Dynll1 and Dynll2) found at residues 1433–1437, 1509–1513 and 1535–1539, respectively. The binding of DLCs to Bassoon is involved in retrograde transport of synaptic components in the axons [34], and could be dependent on Cdk5 dependent modulation. Interestingly, the *C. elegans* Cdk5 homolog inhibits dynein-dependent retrograde trafficking of presynaptic components [35].

The synaptic vesicle cycle is coordinated with a phosphoinositide cycle that regulates several lipid-protein interactions at the presynaptic terminal. The initial stage for this process is the phosphorylation of PIto PI-4 phosphate mediated by PI-4 kinases, which in synaptic vesicles is mediated by PI4KIIα [36]. Interestingly, the predicted Cdk5 sensitive site (Ser51) is located adjacent to the di-leucine motif ($_{56}$ERQPLL$_{61}$) that allows PI4KIIα association with the adapter protein (AP)-3, a sorting receptor that might be relevant to the PI4KIIα targeting of synaptic vesicles [37].

2.6. Proteins involved in dendritic spine remodeling

In addition to its role at the presynaptic terminal, Cdk5 also phosphorylates key substrates at the post-synaptic compartment (Fig. 4). Our analysis suggests new targets that could contribute to the known activities of Cdk5 in the formation and remodeling of dendritic spines, such as TANC proteins and paralemmin-1 (Fig. 4). We identified putative phosphorylation sites in both TANC1 (Ser445) and TANC2 (Ser1480). TANCs are adapter proteins which can bind PSD-95 (Dlg), and that have been proposed to present redundant roles in the maturation of dendritic spines [38]. Likewise, we also identified a putative phosphorylation site in paralemmin-1 (Palm) (Ser116), a lipid-anchored phosphoprotein located at the cytoplasmic face of membranes which plays an essential role in filopodia formation and dendritic spine maturation [39].

2.7. Proteins involved axon development

Several lines of evidence show that Cdk5 is involved in axon development. Amongst the many proteins regulating this process are CTP:phosphocholine cytidylyltransferases (CTTs), key regulatory enzymes for phosphatidylcholine (PC) synthesis. The CTT-B (Pcyt1b) isoform is particularly enriched in brain, where it appears to be responsible for PC synthesis in the distal segment of axons [40]. Accordingly, CTT-B is essential for axonal growth and branching [41]. CTT-B has been shown to be a substrate for Cdk5 phosphorylation both *in vivo* and *in vitro* [41]. Nevertheless the exact phosphorylation site has not been previously reported. Here we identify a putative phosphorylation site located at Ser323, at the C-terminal region. The phosphorylation of this domain has been shown to regulate the association of CTTs with the membrane compartment [42].

The Krüppel type Zinc Finger transcription factor B-cell lymphome (Bcl)-11a/COUP-TP interacting protein (CTIP)-1 is another putative target of Cdk5, which has been shown to participate in the regulation of neuronal morphology. Although the molecular mechanism remains to be clarified, Bcl11a negatively regulates axonal growth and branching through transcriptional regulation of the microtubule-associated

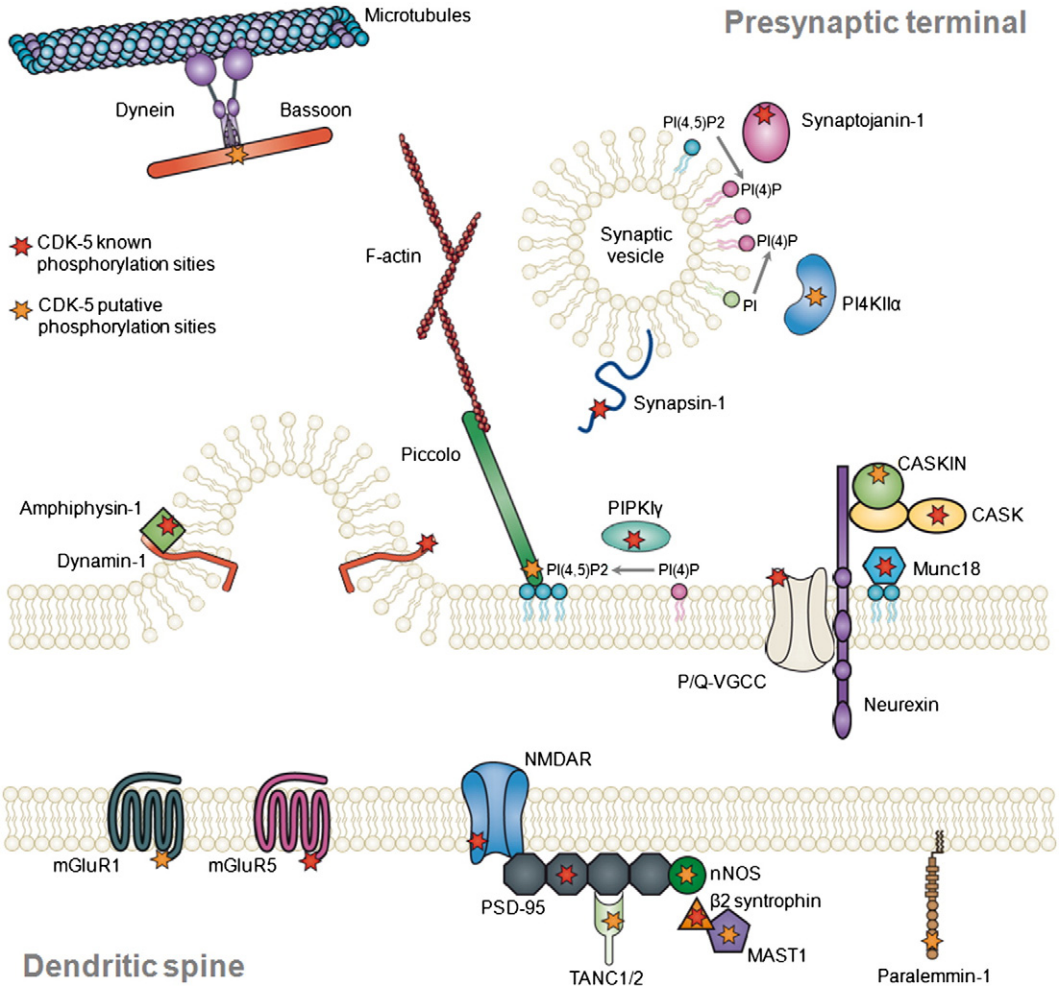

Fig. 4. Known and new putative Cdk5 substrates in the synapses. The contextual information, such as physical interaction, and colocalization or coexpression between kinases and substrates, is essential for the reliability of signal transduction. Interestingly, several of the proteins predicted as putative substrates of Cdk5 in this study interact physically or functionally at synapses with various known substrates of this kinase. CASK, calcium/calmodulin-dependent serine protein kinase; CASKIN, CASK-interacting protein; F-actin, filamentous actin; MAST, microtubule-associated serine-threonine kinase; mGluR, metabotropic glutamate receptor; Munc18, mammalian Unc-18 homolog; NMDAR, N-methyl-D-aspartate receptor; nNOS, neuronal nitric oxide synthase; PI4KIIα, phosphatidylinositol 4-kinase type IIα; PIPKIγ, phosphatidylinositol 4 phosphate 5-kinase type Iγ; PSD-95, postsynaptic density protein 95; TANC, Tetratricopeptide repeat, ankyrin repeat and coiled-coil domain-containing protein; P/Q-VGCC, P/Q-voltage-gated calcium channel.

protein (MAP)-1B and the netrin-1 receptor *D*eleted in *C*olorectal *C*arcinoma (DCC) [43]. Interestingly, the predicted phosphorylation site (Ser332) in CTIP-1 is contained at a SIRT1 binding interface (residues 194-378). SIRT1 is a histone deacetylase shown to be essential for Bcl11a-mediated transcriptional repression [44].

3. Conclusions

We report a novel strategy for proteome-wide identification of previously unreported Cdk5 substrates. Through the comprehensive analysis of the available published data on Cdk5 substrates and their phosphorylation sites, we developed a computational approach that uses a PSSM based search which was complemented with the biological information available from multiple studies identifying phosphorylation sites by mass spectrometry. This allowed us to generate a ranked list of highly likely potential substrates of Cdk5 which can now be subjected to further experimental verification. The highly stringent filtering criteria used results in a high number of false negatives, but permitted the specificity to be maximized. Our strategy presented here may be adapted to identify substrates for other kinases, considering that the only constraint is the availability of enough information (known phosphorylation sites) in order to generate a reliable PSSM.

4. Methods

4.1. Dataset

A comprehensive search was performed of the PubMed database using the keywords “Cdk5” and “Cyclin-dependent kinase 5”. The search yielded 166 experimentally characterized phosphorylation sites in 99 different proteins from 1212 research articles published online before Sept. 30, 2011. The primary sequences for each protein were extracted from the Swiss-Prot database. The dataset was divided into three subgroups according to the experimental determination of the phosphorylation site, accordingly “*in vivo*”, “*in vitro*” and “*in vivo* and *in vitro*”. Although the entire dataset was used to performance tests (ROC curve, Retrieve percentage), a subset of only 96 sites was used for the construction of the PSSM. This corresponds to the dataset named “*in vivo* and *in vitro*” with the exception of cellular tumor antigen p53 (Ser15), which does not correspond to a minimum consensus sequence of Cdk5.

4.2. Construction of a scansite-suitable PSSM and proteome search

96 phosphorylation sites which are physiologically and unambiguously phosphorylated by Cdk5 (*in vivo* and *in vitro* dataset) were

aligned and the apparition frequency (%) of each of 20 amino acids was determined within a fifteen-residue window (eight residues upstream and six residues downstream of the phosphorylation site, Pro). The frequency was normalized to the mean background frequency for each amino-acid in the Swiss-Prot databases of *Mus musculus*, *Homo sapiens* and *Rattus norvegicus*. Thus, the value "1" indicates that an amino acid appearing in a position does so in the same percentage as this amino acid is represented in the proteome. Values lower than 1 indicate that this amino acid is less represented at this position than what would be expected and values above 1 indicate a higher frequency than expected of the amino acid in this position. Values were incorporated into a matrix of 15 rows (positions, with the amino acid Pro fixed in the central position) and 20 columns (amino acids). The developed matrix was incorporated in the option "Database Search > Search using an input motif" of Scansite (scansite.mit.edu) [45]. The search was performed with high-stringency on the Swiss-prot database of *Mus musculus*.

4.3. Prediction of protein disorder

The open access server PrDOS (http://prdos.hgc.jp) [46] was used to predict disordered regions in known substrates of Cdk5 from their amino acid sequence. The analysis considered the template-based prediction option and a false positive rate of 5% which equates to a true positive rate of 65%, according to the ROC curve for the predictor. The server returns a score between 0.5 and 1 if the amino acid is found to be in disordered regions.

4.4. Additional tools

The Pfam database (pfam.sanger.ac.uk) [47] was consulted in order to define the domain architecture of different proteins. Clustal Omega [48] was used to perform multiple alignments and determine residue homology between proteins of different species. PhosphositePlus (www.phosphosite.org) [49], which collects most of the phosphorylation sites identified by mass spectrometry studies, was used to determine whether a site is known to be phosphorylated *in vivo* or not. Only studies using mouse cells or tissues were considered. Additional prediction on Cdk5 substrates was performed using NetPhosK (http://www.cbs.dtu.dk/services/NetPhosK/) [24] and GPS (http://gps.biocuckoo.org/) [25].

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2013.01.003.

Acknowledgments

We thank Dr. Alejandro Roth and Dr. Michael Handford (Universidad de Chile) for manuscript proof-reading and valuable comments. This work was supported by CONICYT Fellowship D-21070173 and 24090187 (to DB), Grant FONDECYT 1095089 (to CG-B) and Grants Basal Project Center for Mathematical Modeling and FONDAP Center for Genome Regulation CGR-15090007 (to AM).

References

[1] F.A. Dhariwala, M.S. Rajadhyaksha, An unusual member of the Cdk family: Cdk5, Cell. Mol. Neurobiol. 28 (2008) 351–369.
[2] E. Contreras-Vallejos, E. Utreras, C. Gonzalez-Billault, Going out of the brain: non-nervous system physiological and pathological functions of Cdk5, Cell. Signal. 24 (2012) 44–52.
[3] J.L. Rosales, K.Y. Lee, Extraneuronal roles of cyclin-dependent kinase 5, Bioessays 28 (2006) 1023–1034.
[4] J.A. Ubersax, J.E. Ferrell Jr., Mechanisms of specificity in protein phosphorylation, Nat. Rev. Mol. Cell Biol. 8 (2007) 530–541.
[5] L.J. Holt, B.B. Tuch, J. Villen, A.D. Johnson, S.P. Gygi, D.O. Morgan, Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution, Science 325 (2009) 1682–1686.
[6] J.A. Ubersax, E.L. Woodbury, P.N. Quang, M. Paraz, J.D. Blethrow, K. Shah, K.M. Shokat, D.O. Morgan, Targets of the cyclin-dependent kinase Cdk1, Nature 425 (2003) 859–864.
[7] E.J. Chang, R. Begum, B.T. Chait, T. Gaasterland, Prediction of cyclin-dependent kinase phosphorylation substrates, PLoS One 2 (2007) e656.
[8] R. Linding, L.J. Jensen, G.J. Ostheimer, M.A. van Vugt, C. Jorgensen, I.M. Miron, F. Diella, K. Colwill, L. Taylor, K. Elder, P. Metalnikov, V. Nguyen, A. Pasculescu, J. Jin, J.G. Park, L.D. Samson, J.R. Woodgett, R.B. Russell, P. Bork, M.B. Yaffe, T. Pawson, Systematic discovery of in vivo phosphorylation networks, Cell 129 (2007) 1415–1426.
[9] T. Sardon, R.A. Pache, A. Stein, H. Molina, I. Vernos, P. Aloy, Uncovering new substrates for Aurora A kinase, EMBO Rep. 11 (2010) 977–984.
[10] U.E. Rennefahrt, S.W. Deacon, S.A. Parker, K. Devarajan, A. Beeser, J. Chernoff, S. Knapp, B.E. Turk, J.R. Peterson, Specificity profiling of Pak kinases allows identification of novel phosphorylation sites, J. Biol. Chem. 282 (2007) 15667–15678.
[11] M.B. Yaffe, G.G. Leparc, J. Lai, T. Obata, S. Volinia, L.C. Cantley, A motif-based profile scanning approach for genome-wide prediction of signaling pathways, Nat. Biotechnol. 19 (2001) 348–353.
[12] H.D. Huang, T.Y. Lee, S.W. Tzeng, J.T. Horng, KinasePhos: a web tool for identifying protein kinase-specific phosphorylation sites, Nucleic Acids Res. 33 (2005) W226–W229.
[13] N. Blom, S. Gammeltoft, S. Brunak, Sequence and structure-based prediction of eukaryotic protein phosphorylation sites, J. Mol. Biol. 294 (1999) 1351–1362.
[14] Y.H. Wong, T.Y. Lee, H.K. Liang, C.M. Huang, T.Y. Wang, Y.H. Yang, C.H. Chu, H.D. Huang, M.T. Ko, J.K. Hwang, KinasePhos 2.0: a web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns, Nucleic Acids Res. 35 (2007) W588–W594.
[15] K. Fujii, G. Zhu, Y. Liu, J. Hallam, L. Chen, J. Herrero, S. Shaw, Kinase peptide specificity: improved determination and relevance to protein phosphorylation, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 13744–13749.
[16] X. Gao, C. Jin, J. Ren, X. Yao, Y. Xue, Proteome-wide prediction of PKA phosphorylation sites in eukaryotic kingdom, Genomics 92 (2008) 457–463.
[17] E.L. Huttlin, M.P. Jedrychowski, J.E. Elias, T. Goswami, R. Rad, S.A. Beausoleil, J. Villen, W. Haas, M.E. Sowa, S.P. Gygi, A tissue-specific atlas of mouse protein phosphorylation and expression, Cell 143 (2010) 1174–1189.
[18] Y. Chi, M. Welcker, A.A. Hizli, J.J. Posakony, R. Aebersold, B.E. Clurman, Identification of CDK2 substrates in human cell lysates, Genome Biol. 9 (2008) R149.
[19] J.D. Blethrow, J.S. Glavy, D.O. Morgan, K.M. Shokat, Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 1442–1447.
[20] J.V. Olsen, B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, M. Mann, Global, in vivo, and site-specific phosphorylation dynamics in signaling networks, Cell 127 (2006) 635–648.
[21] J. Gao, J.J. Thelen, A.K. Dunker, D. Xu, Musite, a tool for global prediction of general and kinase-specific phosphorylation sites, Mol. Cell. Proteomics 9 (2010) 2586–2600.
[22] A.M. Moses, J.K. Heriche, R. Durbin, Clustering of phosphorylation site recognition motifs can be exploited to predict the targets of cyclin-dependent kinase, Genome Biol. 8 (2007) R23.
[23] K. Iijima, K. Ando, S. Takeda, Y. Satoh, T. Seki, S. Itohara, P. Greengard, Y. Kirino, A.C. Nairn, T. Suzuki, Neuron-specific phosphorylation of Alzheimer's beta-amyloid precursor protein by cyclin-dependent kinase 5, J. Neurochem. 75 (2000) 1085–1091.
[24] N. Blom, T. Sicheritz-Ponten, R. Gupta, S. Gammeltoft, S. Brunak, Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence, Proteomics 4 (2004) 1633–1649.
[25] Y. Xue, J. Ren, X. Gao, C. Jin, L. Wen, X. Yao, GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy, Mol. Cell. Proteomics 7 (2008) 1598–1608.
[26] B.A. Joughin, C. Liu, D.A. Lauffenburger, C.W. Hogue, M.B. Yaffe, Protein kinases display minimal interpositional dependence on substrate sequence: potential implications for the evolution of signalling networks, Philos. Trans. R. Soc. Lond. B Biol. Sci. 367 (2012) 2574–2583.
[27] A.M. Hartmann, D. Rujescu, T. Giannakouros, E. Nikolakaki, M. Goedert, E.M. Mandelkow, Q.S. Gao, A. Andreadis, S. Stamm, Regulation of alternative splicing of human tau exon 10 by phosphorylation of splicing factors, Mol. Cell. Neurosci. 18 (2001) 80–90.
[28] J. Shi, T. Zhang, C. Zhou, M.O. Chohan, X. Gu, J. Wegiel, J. Zhou, Y.W. Hwang, K. Iqbal, I. Grundke-Iqbal, C.X. Gong, F. Liu, Increased dosage of Dyrk1A alters alternative splicing factor (ASF)-regulated alternative splicing of tau in Down syndrome, J. Biol. Chem. 283 (2008) 28660–28669.
[29] V.N. Murthy, P. De Camilli, Cell biology of the presynaptic terminal, Annu. Rev. Neurosci. 26 (2003) 701–728.
[30] S.H. Kim, T.A. Ryan, CDK5 serves as a major control point in neurotransmitter release, Neuron 67 (2010) 797–809.
[31] S. Leal-Ortiz, C.L. Waites, R. Terry-Lorenzo, P. Zamorano, E.D. Gundelfinger, C.C. Garner, Piccolo modulation of Synapsin1a dynamics regulates synaptic vesicle exocytosis, J. Cell Biol. 181 (2008) 831–846.
[32] K. Fujimoto, T. Shibasaki, N. Yokoi, Y. Kashima, M. Matsumoto, T. Sasaki, N. Tajima, T. Iwanaga, S. Seino, Piccolo, a Ca2+ sensor in pancreatic beta-cells. Involvement of cAMP-GEFII.Rim2. Piccolo complex in cAMP-dependent exocytosis, J. Biol. Chem. 277 (2002) 50497–50502.
[33] S.H. Gerber, J. Garcia, J. Rizo, T.C. Sudhof, An unusual C(2)-domain in the active-zone protein piccolo: implications for Ca(2+) regulation of neurotransmitter release, EMBO J. 20 (2001) 1605–1619.
[34] A. Fejtova, D. Davydova, F. Bischof, V. Lazarevic, W.D. Altrock, S. Romorini, C. Schone, W. Zuschratter, M.R. Kreutz, C.C. Garner, N.E. Ziv, E.D. Gundelfinger, Dynein light chain regulates axonal trafficking and synaptic levels of Bassoon, J. Cell Biol. 185 (2009) 341–355.
[35] C.Y. Ou, V.Y. Poon, C.I. Maeder, S. Watanabe, E.K. Lehrman, A.K. Fu, M. Park, W.Y. Fu, E.M. Jorgensen, N.Y. Ip, K. Shen, Two cyclin-dependent kinase pathways are

essential for polarized trafficking of presynaptic components, Cell 141 (2010) 846–858.
[36] J. Guo, M.R. Wenk, L. Pellegrini, F. Onofri, F. Benfenati, P. De Camilli, Phosphatidylinositol 4-kinase type IIalpha is responsible for the phosphatidylinositol 4-kinase activity associated with synaptic vesicles, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 3995–4000.
[37] B. Craige, G. Salazar, V. Faundez, Phosphatidylinositol-4-kinase type II alpha contains an AP-3-sorting motif and a kinase domain that are both required for endosome traffic, Mol. Biol. Cell 19 (2008) 1415–1426.
[38] S. Han, J. Nam, Y. Li, S. Kim, S.H. Cho, Y.S. Cho, S.Y. Choi, J. Choi, K. Han, Y. Kim, M. Na, H. Kim, Y.C. Bae, E. Kim, Regulation of dendritic spines, spatial memory, and embryonic development by the TANC family of PSD-95-interacting proteins, J. Neurosci. 30 (2010) 15102–15112.
[39] P. Arstikaitis, C. Gauthier-Campbell, R. Carolina Gutierrez Herrera, K. Huang, J.N. Levinson, T.H. Murphy, M.W. Kilimann, C. Sala, M.A. Colicos, A. El-Husseini, Paralemmin-1, a modulator of filopodia induction is required for spine maturation, Mol. Biol. Cell 19 (2008) 2026–2038.
[40] J. Strakova, L. Demizieux, R.B. Campenot, D.E. Vance, J.E. Vance, Involvement of CTP:phosphocholine cytidylyltransferase-beta2 in axonal phosphatidylcholine synthesis and branching of neurons, Biochim. Biophys. Acta 1811 (2011) 617–625.
[41] J.M. Carter, L. Demizieux, R.B. Campenot, D.E. Vance, J.E. Vance, Phosphatidylcholine biosynthesis via CTP:phosphocholine cytidylyltransferase 2 facilitates neurite outgrowth and branching, J. Biol. Chem. 283 (2008) 202–212.
[42] M.K. Dennis, S.G. Taneva, R.B. Cornell, The intrinsically disordered nuclear localization signal and phosphorylation segments distinguish the membrane affinity of two cytidylyltransferase isoforms, J. Biol. Chem. 286 (2011) 12349–12360.
[43] T.Y. Kuo, C.J. Hong, Y.P. Hsueh, Bcl11A/CTIP1 regulates expression of DCC and MAP1b in control of axon branching and dendrite outgrowth, Mol. Cell. Neurosci. 42 (2009) 195–207.
[44] T. Senawong, V.J. Peterson, M. Leid, BCL11A-dependent recruitment of SIRT1 to a promoter template in mammalian cells results in histone deacetylation and transcriptional repression, Arch. Biochem. Biophys. 434 (2005) 316–325.
[45] J.C. Obenauer, L.C. Cantley, M.B. Yaffe, Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs, Nucleic Acids Res. 31 (2003) 3635–3641.
[46] T. Ishida, K. Kinoshita, PrDOS: prediction of disordered protein regions from amino acid sequence, Nucleic Acids Res. 35 (2007) W460–W464.
[47] M. Punta, P.C. Coggill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E.L. Sonnhammer, S.R. Eddy, A. Bateman, R.D. Finn, The Pfam protein families database, Nucleic Acids Res. 40 (2012) D290–D301.
[48] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, Mol. Syst. Biol. 7 (2011) 539.
[49] P.V. Hornbeck, J.M. Kornhauser, S. Tkachev, B. Zhang, E. Skrzypek, B. Murray, V. Latham, M. Sullivan, PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse, Nucleic Acids Res. 40 (2012) D261–D270.